

Food Demand Forecasting



차례

1. 연구 제목
2. 연구 배경
3. 관련 연구
4. 연구 방법론
5. 기대 효과
6. 참고문헌

1. 연구 제목

회귀분석에 기반한 향후 10주의 식자재 수요 예측

2. 연구 배경

연구 동기

마켓컬리의 “샐러드배송”, 쿠팡의 “로켓프레시”의 성장을 지켜보며 물류 수요 예측에 관심

연구 목적

수요 예측 정확도가 가장 높은 회귀분석 모델을 찾기 위함

연구 문제

Pycaret Regression은 수요 예측 모델의 정확도에 영향을 미칠까?

3. 관련 연구

기존 연구

- 회귀분석 모델을 임의로 선택
- 해당 모델의 파라미터를 조정
 - RandomForestRegressor
 - LinearRegression
 - DecisionTreeRegressor

기존 연구와의 차별성

- Pycaret Regression
- 회귀 모델 간의 결과 비교
- 최적화 된 파라미터 값 도출
- 파라미터 튜닝을 통해 모델링
 - RandomSearch
 - GridSearch

4. 연구 방법론

데이터 설명

- Kaggle의 "Food Demand Forecasting" 식자재 물류창고 데이터
- 데이터 수집시기는 알 수 없으나, 약 2년 전에 데이터가 업데이트 됨
- 특정 week부터 약 145주간의 정보 제공
- Train은 9개의 column과 456,548개의 row로 구성
 - 기본적으로 "train", "test" 및 "sample_submission" 데이터 제공
 - 변수 설명을 위한 "fulfilment_center_info", "meal_info" 추가 제공

4. 연구 방법론

데이터 변수 설명

- Index : 인덱스
- Id : 고유번호
- Week : 1주~145주까지 나타냄
- Center_id : 각 도시에 있는 물류창고 번호
- Meal_id : 음식 품목과 종류
- Checkout_price : 구매가격
- Base_price : 원가격
- Emailer_for_promotion : 이메일을 통한 프로모션 유무
- Homepage_featured : 홈페이지 게재 유무
- Num_orders : 주문 개수

Index	Id	week	center_id	meal_id	checkout_price	base_price	emailer_for_promotion	homepage_featured	Num_orders
0	1379560	1	55	1885	136.83	152.29	0	0	177
1	1466964	1	55	1993	136.83	135.83	0	0	270
2	1346989	1	55	2539	134.86	135.86	0	0	189
3	1338232	1	55	2139	339.5	437.53	0	0	54
4	1448490	1	55	2631	243.5	242.5	0	0	40
...
456543	1271326	145	61	1543	484.09	484.09	0	0	68
456544	1062036	145	61	2304	482.09	482.09	0	0	42
456545	1110849	145	61	2664	237.68	321.07	0	0	501
456546	1147725	145	61	2569	243.5	313.34	0	0	729
456547	1361984	145	61	2490	292.03	290.03	0	0	162

	center_id	city_code	region_code	center_type	op_area
0	11	679	56	TYPE_A	3.7
1	13	590	56	TYPE_B	6.7
2	124	590	56	TYPE_C	4.0
3	66	648	34	TYPE_A	4.1
4	94	632	34	TYPE_C	3.6

Fulfilment_center_info

	meal_id	category	cuisine
0	1885	Beverages	Thai
1	1993	Beverages	Thai
2	2539	Beverages	Thai
3	1248	Beverages	Indian
4	2631	Beverages	Indian

Meal_info

4. 연구 방법론

활용 가능한 마이닝 기법

- 회귀분석
 - PycaretRegressor
 - RandomForestRegressor
 - DecisionTreeRegressor
 - CatboostRegressor 등
 - Hyper Parameter
 - GridSearch
 - RandomSearch

```
# 다양한 모델들과의 성능 비교  
compare_models()
```

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE	TT (Sec)
catboost	CatBoost Regressor	2.3579	41.8379	6.2979	0.0796	0.4287	0.4397	1.8100
gbr	Gradient Boosting Regressor	2.3877	42.9870	6.3553	0.0794	0.4304	0.4498	1.2920
lightgbm	Light Gradient Boosting Machine	2.4064	42.7072	6.3550	0.0700	0.4405	0.4537	0.1220
ridge	Ridge Regression	2.4465	46.9089	6.6109	0.0182	0.4429	0.4621	0.0390
lar	Least Angle Regression	2.4498	47.0098	6.6182	0.0159	0.4437	0.4627	0.0450
br	Bayesian Ridge	2.4593	47.3046	6.6309	0.0156	0.4408	0.4661	0.2960
omp	Orthogonal Matching Pursuit	2.4628	47.2961	6.6402	0.0089	0.4436	0.4656	0.0420
lasso	Lasso Regression	2.4768	47.8551	6.6703	0.0037	0.4449	0.4710	0.7460
en	Elastic Net	2.4768	47.8551	6.6703	0.0037	0.4449	0.4710	0.7760
lr	Linear Regression	2.4768	47.8551	6.6703	0.0037	0.4449	0.4710	0.4330
llar	Lasso Least Angle Regression	2.4785	48.0680	6.6857	-0.0010	0.4458	0.4718	0.0420
dummy	Dummy Regressor	2.4785	48.0680	6.6857	-0.0010	0.4458	0.4718	0.0340
huber	Huber Regressor	2.0249	49.8582	6.8198	-0.0447	0.4453	0.2681	0.4270
ada	AdaBoost Regressor	2.9807	49.0434	6.7335	-0.0483	0.4976	0.6459	0.3870
knn	K Neighbors Regressor	2.5518	47.6331	6.7733	-0.0817	0.4852	0.4653	0.3420
rf	Random Forest Regressor	2.5254	47.0310	6.7306	-0.0930	0.4712	0.4664	3.0900
et	Extra Trees Regressor	2.5966	53.9908	7.2176	-0.2667	0.4861	0.4773	3.5160
dt	Decision Tree Regressor	2.6340	56.1225	7.3445	-0.3015	0.4933	0.4814	0.0920
par	Passive Aggressive Regressor	16.9048	1053.1827	20.2289	-41.6919	1.0367	4.4461	0.0790

```
<catboost.core.CatBoostRegressor at 0x19eef11c848>
```


5. 기대 효과

- **신선식품 폐기율 감소를 통한 비용 절감 및 환경보호**
- **식품 관리 역량을 강화하여 효율적인 물류 시스템 구축**
- **물류창고의 인력을 효율적으로 배치하여 비용 절감**
- **식품 수요를 예측하여 상품 발주**

6. 참고문헌

- 데이터 출처 : Kaggle, Food Demand Forecasting.
<https://www.kaggle.com/datasets/kannanaikkal/food-demand-forecasting>
- 김하루, *마켓컬리 7년간 식품 폐기율 1%유지...빅데이터 기반 예측*, 식품외식경영, 20220311,
<http://www.foodnews.news/mobile/article.html?no=475458>
- 박세준, *전통 모빌리티 수요 예측 정확도 개선* (한국정보기술학회, 2021), 104-106
- 오지연, *라스트마일 물류의 빅데이터 활용 사례 분석* (한국정보통신학회 종합학술대회 논문집, 2019), 121-123
- 정선우, *"내일은 얼마나 팔릴까?" 외식 수요예측 고민 시로 해결한다*, 푸드경제신문, 20220224,
<http://www.foodneconomy.com/news/articleView.html?idxno=333669>

감사합니다

