

# 코르카 모델 내제화 진척상황(10/18일기준)

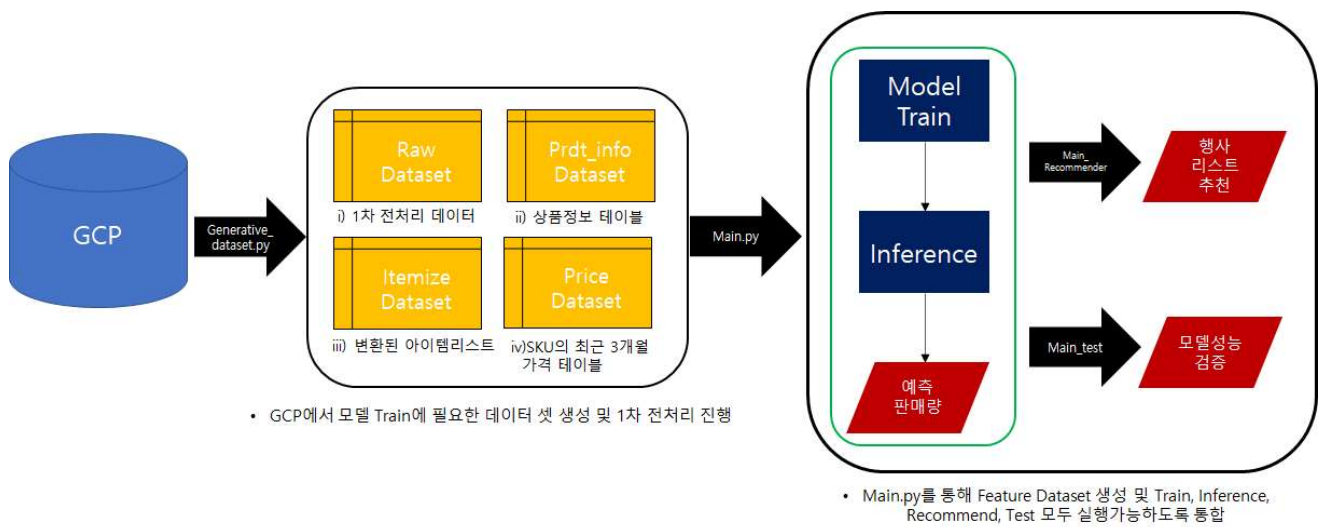
(이전까지 진척상황: [\[21.09.28\] MIC Bi-weekly\\*](#))

\*Corca Model: MGS recommendor v2.0

Steps: 1) Neural ODE(상품의 판매량 예측) → 2) MGS recommender(조건을 부합하는 행사리스트의 SKU Decision) → 행사유형, 진열위치를 고려한 행사상품 List 선정

코르카 모델 내제화 단계	Python 코드	완료여부
1)GCP→ raw_dataset <ul style="list-style-type: none"><li>Raw Dataset( 1차 전처리 데이터)</li><li>Prdt Info Dataset(SKU의 상세정보)</li><li>Itemize Dataset(SKU단위→ Item 단위로 통합한 데이터셋)</li><li>Price Dataset(최근 3개월 상품가격)</li></ul>	Generative_dataset.py	O
2)raw→ Feature dataset <ul style="list-style-type: none"><li>Model Training에 사용하는 실제 Feature 셋</li><li>3,6,12주간 가격 평균 및 최댓값등의 변수 생성</li></ul>	*Main.py <ul style="list-style-type: none"><li>utils.py 부분</li></ul>	O
3)Feature dataset→ Model Train <ul style="list-style-type: none"><li>대분류별 2개의 모델 생성(판매량 기준 Top200, Top 3000)</li><li>pkl파일로 각 모델 저장</li></ul>	*Main.py <ul style="list-style-type: none"><li>action: Train으로 설정</li></ul>	O
4)Model Train→ Model Inference(판매량예측)	*Main.py <ul style="list-style-type: none"><li>action: Inference로 설정</li></ul>	O
5) Model Inference→ MGS_Recommendor(조합식 극대화를 통해 행사리스트 추천) <ul style="list-style-type: none"><li>코르카에서 제안하는 아웃풋과 동일함</li></ul>	*Main.py <ul style="list-style-type: none"><li>action: recommendor로 설정</li></ul>	O
6) Model inference→ Model Test(예측판매량 모델과 실제 판매량 모델 비교를 통해 성능검증) <ul style="list-style-type: none"><li>MAPE, R2 Score등의 지표 추가</li><li>Tensor Board 시각화</li><li>모델별 output을 바탕으로 성능비교</li></ul>	*Main.py <ul style="list-style-type: none"><li>actinon: valid로 설정</li></ul>	△
7) 코드 효율화 구동시간 개선(Python 코드 기준) <ul style="list-style-type: none"><li>각 Step 별 log 생성 및 Config 파일 기록</li><li>각 STEP별 Config 파일을 통해 코드 자동화</li></ul> <p>*코드 효율화작업이 어느정도 진행되었지만, 효율화 및 개선방향 고민후 적용예정</p>	*Main.py	△
8) 모델 고도화 <ul style="list-style-type: none"><li>파라미터 최적화</li><li>변수 추가</li></ul>		X

\*내제화 모델 Flow



### 1) Generative Dataset.py

- config 파일에 기간, 점포, 대분류코드를 입력하면 자동으로 GCP상에서 데이터 loading 후 자동으로 4개의 데이터셋 생성

-Raw\_Dataset, Prdt\_mastr\_Dataset, Itemize\_Dataset, Price\_Dataset등 생성

```

▶python generate_dataset v4.0.py --config config.json
-----
데이터셋 시작일자: 2020-11-05
데이터셋 종료일자: 2021-02-10
대상점포 코드: 1024,1117,1084
대상대분류 코드: 310,460,320
담당 코드: 30,40
itemize:False
-----
아이템 리스트파일이 존재합니다
RCP_MASTR_TA 데이터 로딩을 완료하였습니다.
target_normal_event,target_pos_event 데이터 로딩을 완료하였습니다.
최근3개월 가격의 시작일자: 2021-07-17
최근3개월 가격의 끝일자: 2021-10-14
*****
데이터셋 변환을 완료하였습니다.
3053it [00:04, 700.55it/s]
행사유형_1117점 포변환완료
3100it [00:04, 695.72it/s]
행사유형_1084점 포변환완료
3053it [00:04, 701.68it/s]
행사유형_1024점 포변환완료
254it [00:00, 658.71it/s]
1+1,2+1_1117점 포변환완료
254it [00:00, 660.04it/s]
1+1,2+1_1084점 포변환완료
254it [00:00, 652.73it/s]
1+1,2+1_1024점 포변환완료
Price dataset이 없으므로 생성합니다
*****
Raw Dataset 저장을 완료하였습니다.
*****
진행된 데이터 관련 정보를 저장합니다.
*****
진행된 데이터 관련 정보를 저장 경로: ./stage1/raw_info_1634285956.json
*****
모델구동시간: 0:00:40

```

#### i) GCP → Raw Dataset

- 일자별-점포-상품의 판매정보, 행사유형정보등이 기록된 테이블

	STR_CD	PRDT_CD	BSN_DT	PRDT_NM	PRDT_GCODE_CD	PRDT_MCODE_CD	PRDT_DCODE_CD	NET	SALE_QTY	UCOST_AMT	...	포 리 표 출 력	상 품 에 누 리
0	1024	49181524	2020-11-05	S&B 유즈 코소 80g	310	3763	118	7980	1	4933	...	0	0
1	1024	49181524	2020-11-06	S&B 유즈 코소 80g	310	3763	118	0	0	4933	...	0	0
2	1024	49181524	2020-11-07	S&B 유즈 코소 80g	310	3763	118	0	0	4933	...	0	0
3	1024	49181524	2020-11-08	S&B 유즈 코소 80g	310	3763	118	0	0	4933	...	0	0

#### ii) GCP→PRDT\_MAST\_INFO

- SKU의 정보 테이블

	PRDT_CD	PRDT_NM	PRDT_GCODE_CD	PRDT_MCODE_CD	PRDT_DCODE_CD	PRDT_GCODE_NM	PRDT_MCODE_NM	PRDT_DCODE_NM
0	8802867112085	최초장 2.2k	310	3760.0	6.0	조미료	고추장	초고추장
1	8801558600214	음트리히 조장13kg	310	3760.0	6.0	조미료	고추장	초고추장
2	8801390125531	해)새콤초 고추장320	310	3760.0	6.0	조미료	고추장	초고추장
3	8801007053752	쇠고기비빔고추장 340G	310	3760.0	6.0	조미료	고추장	초고추장

### iii)GCP→ ITEMIZE

- 코르카에서 개발된 로직에 따라 SKU→ ITEM으로 변환하여 생성한 테이블  
\*실제 Recommendor 사용시 매입본부에서 Conform한 아이템 리스트를 활용함

	ITEM	PRDT_NM	PRDT_CD	PRDT_GCODE_CD	URPRICE
0	1	청정원 매운고추장 1kg	8801052802046	310	15300
1	1	청정원 태양초참고추장 1kg	8801052435022	310	15300
2	2	청정원 만능 비빔장 300g	8801052025209	310	4300
3	2	청정원 매콤만능볶음장 300g	8801052025193	310	4300
4	3	올반키친 쇠고기볶음고추장소스 180g	8809547911426	310	6900

### iv)GCP→ PRICE Dataset

- SKU 별로 최근 3개월치 상품 가격이 기록된 테이블  
\*원가 변경이슈로 Inference시 최근 3개월의 데이터 활용

	PRDT_CD	UCOST_AMT	UPRICE_AMT	ENURI_AMT
0	8801007406626	2973.0	5000.0	0.000000
1	8809380343538	1511.0	2480.0	0.000000
2	8801045427102	2085.0	3450.0	0.000000
3	8801047416487	2196.0	3980.0	0.000000
4	8809112895465	2400.0	3480.0	0.000000

## 2)Main.py

-Generataive Dataset.py에서 생성된 Ouput 및 log 기록을 받아서 상품판매량 예측모델 (Neural ODE)을 통해 Train, Inference, Test, recommendation 등을 진행

- action type의 변경을 통해 해당유형을 변경 Ex) Train의 경우 log파일에 "action": "train"을 추가 입력

### i) Main.Train(Model Train 부분)

- 대분류 코드별로 모델학습 진행. Sku\_split 옵션을 통해 판매량이 높은 SKU기준 Top 200, Top 3000일 때의 Case를 나누어 2개의 모델 학습후 저장(SKU\_Split은 변경가능)

```
python main v5.py --config ./stage1/raw_info 1634279384.json
```

```
-----
데이터셋 시작일자: 2020-11-05
데이터셋 종료일자: 2021-02-10
대상점포 코드: ['1024', '1117', '1084']
대상대분류 코드: 310,460,320
학습유형 (action type): train
데이터마트사용여부: False
daily_path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/raw/rcipt_daily_1634279384.csv
prdt_info_path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/raw/prdt_mastr_info_1634279384.csv
price_path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/raw/prdt_price_1634279384.csv
item_path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/files/item.xlsx
pred_path: ./result/
valid_path: ./result
-----
```

```
/workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/raw/rcipt_daily_1634279384.csv
```

```
Filtering gcodes [310]
```

```
end init...
```

```
MGS : Training [310] & 50 item models
```

```
Processing '_train_idx'.
```

```
top k 50
```

epoch	r2	train_loss	valid_loss	dur
1	0.4170	1.3190	3.3155	0.0151
2	-0.3273	3.0029	1.2423	0.0064
3	0.4725	1.1933	1.2530	0.0057
4	0.4982	1.1353	1.2845	0.0056
5	0.4577	1.2269	1.0629	0.0059
6	0.5135	1.1007	0.8681	0.0057

## ii) Main.inference(Model inference 부분)

- Train을 완료한 Model을 load하여 각 대분류별로 inference 진행
- Train 후 생성된 model\_info\_log 파일에서 action을 train→inference로 변경



```

▶ python main v5.py --config ./stage2/model_info 1634279632.json
-----
데이터셋 시작일자: 2020-11-05
데이터셋 종료일자: 2021-02-10
대상점포 코드: ['1024', '1117', '1084']
대상대분류 코드: 310,460,320
학습유형(action type): inference
데이터마트사용여부: False
daily_path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/raw/rcipt_daily_1634279384.csv
prdt_info_path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/raw/prdt_mastr_info_1634279384.csv
price_path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/raw/prdt_price_1634279384.csv
item_path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/files/item.xlsx
pred_path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/result/mgs_pred_1634279536.csv
valid_path: ./result
-----
/workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/raw/rcipt_daily_1634279384.csv
Filtering gcodes [310]
end init...
Processing 'inference_idx'.
/workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/result/model_log_1634279536.csv
top_k 50
Inference Done.
Processing 'inference_idx'.
/workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/result/model_log_1634279536.csv
top_k 300
Inference Done.
/workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/raw/rcipt_daily_1634279384.csv
Filtering gcodes [460]
end init...
Processing 'inference_idx'.
/workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/result/model_log_1634279536.csv
top_k 50
Inference Done.
Processing 'inference_idx'.
/workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/result/model_log_1634279536.csv
top_k 300
Inference Done.
/workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/raw/rcipt_daily_1634279384.csv
Filtering gcodes [320]
end init...
Processing 'inference_idx'.
/workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/result/model_log_1634279536.csv
top_k 50
Inference Done.
Processing 'inference_idx'.
/workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/result/model_log_1634279536.csv
top_k 300
Inference Done.

```

- Inference 결과
- y: 예측판매량, Price: 최근 3개월 SKU가격의 평균, EVENT\_POS: 진열매대 위치, EVENT\_TYPE: 행사유형

	y	PRICE	BSN_DT	PRDT_CD	STR_CD	EVENT_POS	EVENT_TYPE
0	27.689026	15899.999290	2021-02-24	8801005111003	1024	ec	0
1	35.516780	14309.899471	2021-02-24	8801005111003	1024	ec	10
2	46.050324	12719.799098	2021-02-24	8801005111003	1024	ec	20
3	61.966265	11129.699886	2021-02-24	8801005111003	1024	ec	30
4	88.201014	9539.599477	2021-02-24	8801005111003	1024	ec	40

### iii) Model Test

- Neural Ode로 예측한 판매량과 실제 판매량을 비교하여 모델 성능 검증
- 모델의 예측시점(1주일단위)에 해당하는 기간을 gcp 상에서 쿼리로 불러와서 비교함  
(일주일간 점포-상품-행사유형-매대의 실제 판매량 vs 아웃풋 기준 점포-상품-행사유형-매대의 예측판매량)
- inference 후 생성된 model\_info\_log 파일에서 action을 inference → valid 로 변경하여 사용

```

▶ python main_v5.py --config ./stage2/model_info 1634279632.json
-----
데이터셋 시작일자: 2020-11-05
데이터셋 종료일자: 2021-02-10
대상점포 코드: ['1024', '1117', '1084']
대상대분류 코드: 310,460,320
학습유형 (action_type): valid
데이터마트사용여부: False
daily_path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/raw/rcipt_daily_1634279384.csv
prdt_info_path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/raw/prdt_mastr_info_1634279384.csv
price_path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/raw/prdt_price_1634279384.csv
item_path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/files/item.xlsx
pred_path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/result/mgs_pred_1634279536.csv
valid_path: ./result
-----
/workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/raw/rcipt_daily_1634279384.csv
Filtering gcodes [310, 460, 320]
end init...
아이템 리스트파일이 존재합니다
RCP_MASTR_TA 데이터 로딩을 완료하였습니다.
target normal event,target_pos_event 데이터 로딩을 완료하였습니다.
최근3개월 가격의 시작일자: 2021-07-17
최근3개월 가격의 끝일자: 2021-10-14
690it [00:00, 2941.03it/s]
행사유형_1084점 포변환완료
690it [00:00, 2987.00it/s]
행사유형_1024점 포변환완료
690it [00:00, 3011.01it/s]
행사유형_1117점 포변환완료
162it [00:00, 2340.38it/s]
1+1,2+1_1084점 포변환완료
162it [00:00, 2348.18it/s]
1+1,2+1_1024점 포변환완료
162it [00:00, 2348.78it/s]
1+1,2+1_1117점 포변환완료
*****
validation_dataset 생성을 완료하였습니다.
*****
MAPE Score: 813.0561084448441
r2_score: 0.5537892594798267
*****
진행된 데이터 관련 정보를 저장 경로: ./stage2/model_info_1634287369.json
*****
모델 구동시간: 0:00:11

```

#### iv) Recommender Model

- Neural Ode에서 계산된 판매량을 바탕으로, 상품 매출액, 이익액등의 조합을 고려하여 대분류 별로 최적의 행사 리스트 추천
- inference 후 생성된 model\_info\_log 파일에서 action을 inference → recommend 로 변경하여 사용

```
python main v5.py --config ./stage2/model_info_1634279632.json
-----
데이터셋 시작일자 : 2020-11-05
데이터셋 종료일자 : 2021-02-10
대상제품 코드 : ['1024', '1117', '1084']
대상대분류 코드 : 310,460,320
측정유형 (action type): recommend
데이타마트사용여부 : False
daily_path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/raw/rcipt_daily_1634279384.csv
prdt_info path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/raw/prdt_mastr_info_1634279384.csv
price_path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/raw/prdt_price_1634279384.csv
item_path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/files/item.xlsx
pred path: /workspace/dev-corca/wonwuk.lee/MGS_xmas_develop/MGS_xmas_v2.0/result/mgs_pred_1634279536.csv
valid path: ./result
-----
Filtering gcodes [310, 460, 320]
There are 100 options
100%|██████████████████████████████████████████████████████████████████████████| 18/18 [00:00<00:00, 1131.15it/s]
(137, 576)
Number of products = 18
Problem solved in 52 milliseconds
Problem solved in 18 iterations
Problem solved in 1 branch-and-bound nodes
item 8801005111003
item 8801052079721
item 8801045220314
item 8801039912843
There are 100 options
100%|██████████████████████████████████████████████████████████████████████████| 37/37 [00:00<00:00, 1073.05it/s]
(175, 1184)
Number of products = 37
Problem solved in 121 milliseconds
Problem solved in 48 iterations
Problem solved in 1 branch-and-bound nodes
item 8800172924884
item 8809299233609
item 8801260519262
item 8801166051446
There are 100 options
100%|██████████████████████████████████████████████████████████████████████████| 25/25 [00:00<00:00, 1079.79it/s]
(151, 800)
Number of products = 25
Problem solved in 77 milliseconds
Problem solved in 30 iterations
Problem solved in 1 branch-and-bound nodes
item 8801085069713
item 8801045526126
item 8801354401176
item 8809296884057
MGS Recommender 모델구동완료하였습니다.
```

### \*\*\*최종 행사리스트 제안 Output

- 대분류별로 제약식에 맞춰서 행사유형-진열매대 위치를 고려하여 상품 리스트 출력

[illegible]



**\*진행순서 예시**

- 1) Raw Dataset 생성: python generate\_dataset\_v4.0.py --config config.json (config파일에, 기간, 점포, 대분류코드 등 작성)
- 2) model train: python main\_v5.py --config ./stage1/raw\_info\_1634279384.json (raw data set log 파일 기준, action: "train" 추가)
- 3) model inference: python main\_v5.py --config ./stage2/model\_info\_1634279536.json (train 후 쌓인 log 파일 기준 action: "train" -> "inference" 변경)
- 4) model valid: python main\_v5.py --config ./stage2/model\_info\_1634279632.json (inference 후 쌓인 log 파일 기준 action: "inference" -> "valid" 변경)

**\*코르카 모델 및 모듈화 개선사항 (개선할 때마다 추가로 기록할 예정)**

- 1) 점포코드, 대분류 코드 확장
- 2) 모델에 사용한 데이터 (Raw Data set, Price Data Set 등) local CSV 파일 → GCP에서 불러올 수 있도록 변환
- 3) 각자 분리되어 있던 파일 로딩, 전처리 과정 → 하나의 Class로 통합 후 자동화
- 5) 코드 효율화 및 구동시간 개선 (Python 코드 기준)
- 4) 모델 고도화
  - 모델 평가 지표 추가 (Validation 검증)
  - 파라미터 최적화 작업 진행
  - Tensorboard 시각화 및 모델 옵션마다 성능 평가 (진행 예정)