

 LG전자 와 함께하는

제 3 회
데이터 사이언스
경진대회
10.27-12.31

POSTECH
UNIST
KAIST

공동주최

주관

 포항공과대학교
산업경영공학부

후원

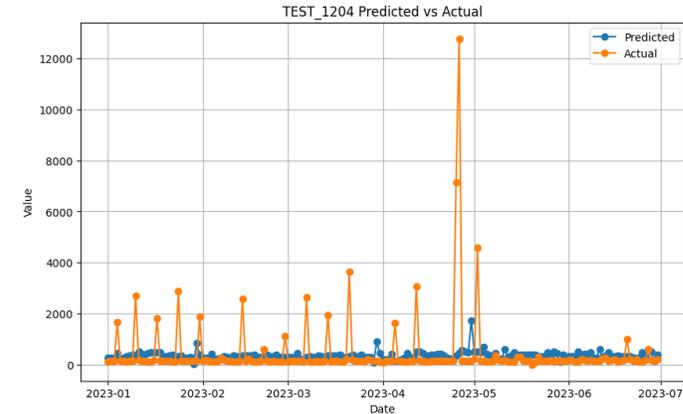
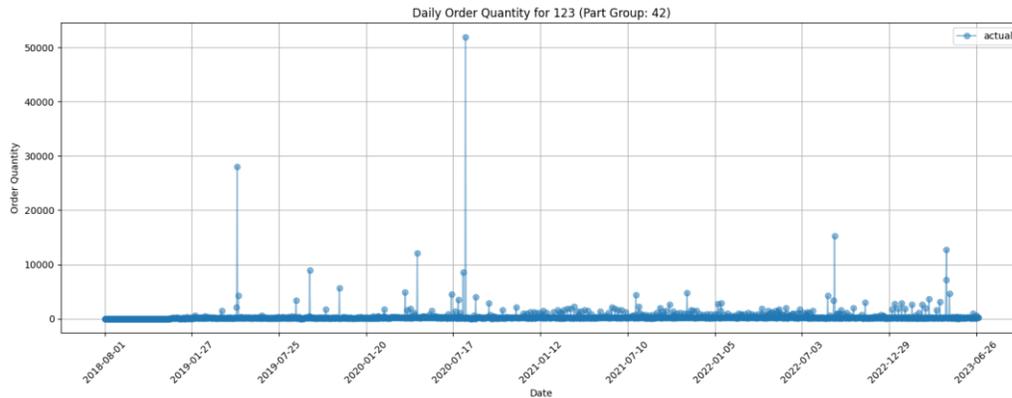
 LG전자  ECMiner  Hankook
& Company  POSTECH
HOLDINGS  PuzzleData

목차

1. 대회 소개
2. 문제 정의
3. 제공 데이터
4. 세부 규칙

대회 소개

- 산업공학: 산업과 사회 시스템 문제해결을 위한 최적 의사결정 + 고성능 예측모델 = 데이터 사이언스



(예비)산업공학도들 모두 함께,
데이터사이언스로 산업/사회의 문제를 해결해보자!
올해는 전자회사의 서비스 부품 수요예측 및 최적화!

대회 소개

• 참가 조건

- POSTECH, UNIST, KAIST의 학부생들로 구성된 팀으로 참가
- 팀 구성 조건 : 4인 이하, 무학과(새내기) 학생 1명 포함 필수,
- 산공과 학부생 1명 포함 필수, 타전공 학생 포함 가능, 서로 다른 학교 학생들로 구성 가능
- 팀 구성이 어려울 시 개인 자격으로 신청 가능 (주최측에서 팀 매칭)



대회 소개

- 대회 일정

- <https://www.datascience-contest.com/>내의 구글폼으로 신청

접수

10.27. ~ 11.10.

대회일정

10.27. ~ 12.31.

문제 공개 : 10.27 (금)

예선작 중간 제출 마감 : 11.25 (토)

예선작 중간 평가 : 12.1 (금)

예선작 최종 제출 마감 : 12.16 (토)

예선작 최종 평가 : 12.21 (목)

본선 진출팀 최종 평가발표 : 12.28 (목)

수상작 시상 : 24년 1월 초

대회 일정은 상황에 따라 일부 변동될 수 있음



대회 소개

• 제출 사항

- 제출물: 1) 문제 해결에 사용한 코드 (.py, .ipynb, R 형식 등), 2) 양식에 맞는 csv 결과 파일, 3) 발표자료 (본선 진출팀에 한함)
- 중간 제출기한: 2023년 11월 25일
- 최종 제출기한: 2023년 12월 16일
- 제출처:
 - 김종원 조교 (pioneer0517@postech.ac.kr)
 - 고영명 교수 (youngko@postech.ac.kr)

문제 공개 : 10.27 (금)
예선작 중간 제출 마감 : 11.25 (토)
예선작 중간 평가 : 12.1 (금)
예선작 최종 제출 마감 : 12.16 (토)
예선작 최종 평가 : 12.21 (목)
본선 진출팀 최종 평가 발표 : 12.28 (목)
수상작 시상 : 24년 1월 초
대회 일정은 상황에 따라 일부 변동될 수 있음

대회 소개

• 대회 상금 및 후원사 안내

- 대상 (1팀): 300만원
 - (LG전자상)
- 금상 (3팀): 200만원
 - (ECMiner상, 포스텍 홀딩스상, 한국엔 컴퍼니상)
- 은상 (1팀): 100만원
 - (퍼즐데이터상)
- 동상 (1팀): 50만원
 - (퍼즐데이터상)
- 장려상 (5팀 내외): 상품 제공

후원사에 대한 자세한 소개는

<https://www.datascience-contest.com/> 참고



- 과제 명: 서비스 부품 수요예측 모델 구축

- 문제 상황

- 서비스 부품의 재고 부족/과잉이 발생할 경우 제반 비용 상승 및 고객만족도 하락으로 이루어짐

- 과제 목적 및 목표

- 서비스 부품의 미래수요를 정확하게 예측하여 서비스 지연에 따른 고객 만족도 상승
- 재고 부족/과잉으로 인해 야기되는 비용을 고려하여 서비스 부품의 공급량을 관리하여 최적의 비용 시스템 구축

- 훈련 데이터: 2018/07/27 (금) – 2023/08/04 (금)

- 경진대회 데이터: 2023/07/31 (월) – 2023/10/29 (일)

- 과제 1 서비스 부품 수요예측 모델 구축

- 부품별 주간 주문량을 예측

- 주간 주문이 없는 경우를 포함하여 부품별 주간 주문량 예측

- 2018/07/27 – 2023/08/04 데이터를 통해 **3개월 단위**의 주간 예측 모델 학습

- 2023년 08월 04일 이전에 관측된 외부데이터 활용 가능

part_no	year	week	Task 1	start_date	end_date
TEST_1045	2023	26		2023-06-26	2023-07-02
TEST_1045	2023	27		2023-07-03	2023-07-09
TEST_1045	2023	28		2023-07-10	2023-07-16
TEST_1045	2023	29		2023-07-17	2023-07-23
TEST_1045	2023	30		2023-07-24	2023-07-30
TEST_1045	2023	31		2023-07-31	2023-08-06
TEST_1045	2023	32		2023-08-07	2023-08-13
TEST_1045	2023	33		2023-08-14	2023-08-20
TEST_1045	2023	34		2023-08-21	2023-08-27
TEST_1045	2023	35		2023-08-28	2023-09-03
TEST_1045	2023	36		2023-09-04	2023-09-10
TEST_1045	2023	37		2023-09-11	2023-09-17

- 주간데이터는 월요일부터 일요일까지 데이터를 합산

- 연도는 속한 날짜 비율에 따름 (isocalendar 함수 사용하여 바로 계산 가능)

- Ex) 2019-12-30(월) 부터 2020-01-05(일)은 2020년 1주차로 명명됨.

- 2020-12-28(월) 부터 2021-01-03(일)은 2020년 53주차로 명명됨.

- 과제 1 서비스 부품 수요예측 모델 구축

- 최종 목적식

- 선정된 부품 10개 relative MSE의 평균 계산
- *relative MSE: 모델의 MSE/ zero-constant function의 MSE
- *zero-constant function: 모든 주간의 수요를 0으로 예측하는 모델

- 선정된 부품 10개

- 'TEST_145', 'TEST_4290', 'TEST_493', 'TEST_141', 'TEST_779', 'TEST_3569', 'TEST_4011', 'TEST_523', 'TEST_1045', 'TEST_1464'
- 주문이 많은 경우 (6개): 'TEST_145', 'TEST_4290', 'TEST_493', 'TEST_141', 'TEST_779', 'TEST_3569'
- 주문이 적당한 경우 (3개): 'TEST_4011', 'TEST_523', 'TEST_1045',
- 주문이 적은 경우 (1개): 'TEST_1464'

• 과제 2 비용을 고려한 서비스 부품 수요예측 모델 구축

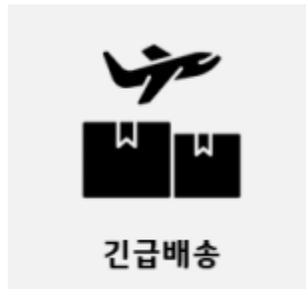
- 비용을 고려하여 부품별 주간 주문수를 예측
- 주어진 비용

- Shipping order cost, $cost_{shipping}$: 주문 부품 개수마다 발생하는 비용
- Airplane order cost, $cost_{airplane}$: 주문 수보다 재고가 부족한 1개당 발생하는 비용
- Carrying cost, $cost_{carrying}$: 1주일마다 재고 1개당 발생하는 비용

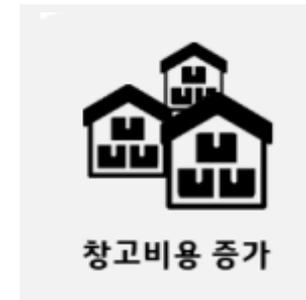
주문 개수당 발생



재고 부족시 발생



재고 과잉시 발생



• 과제 2 비용을 고려한 서비스 부품 수요예측 모델 구축

• 최종 목적식

- 과제1 과 동일하게 선정된 부품 10개에 해당하는 relative total cost의 평균 계산
- *relative total cost: 모델의 total cost/zero-constant function의 total cost

• 부품별 비용(total cost) 계산식

- Total cost = $\sum_{t \in \text{test period}} Cost_t$
- $Cost_t = cost_{carrying} * I_t + cost_{shipping} * y_t^{pred} + cost_{airplane} * \max((y_t^{true} - y_t^{pred} - I_t, 0))$
- $I_{t+1} = \max(I_t + y_t^{pred} - y_t^{true}, 0), I_0 = 0$
- I_t : inventory level at t , y_t^{true} : true weekly order at t , y_t^{pred} : predicted weekly order at t

• 과제 1과 동일한 제출양식

- 과제별 점수 계산
 - 부품별 점수의 평균값
 - part_no: 선정된 10개의 부품 id
- 최종 점수 계산
 - 최종 점수 계산 방식은 추후 공지

점수 계산 예시
(2023-01-01 - 2023-06-30에 대한 예측 결과)

part_no	Model 1	
	과제 1 점수	과제 2 점수
TEST_145	0.295454	0.288102
TEST_4290	0.40393	0.324978
TEST_493	0.124205	0.316121
TEST_141	0.356072	0.308258
TEST_779	0.872964	0.625297
TEST_3569	0.395122	0.403404
TEST_4011	0.370518	0.337317
TEST_523	0.87381	0.523928
TEST_1045	0.251747	0.497034
TEST_1464	1.375	0.708452

모델명	과제 1 점수	과제 2 점수	최종 점수
Model 1	0.53188	0.43329	추후 공지

훈련 데이터: 2018/07/27 (금) – 2023/08/04 (금)
경진대회 데이터: 2023/07/31 (월) – 2023/10/29 (일)

1. 일간 데이터 (327743 rows 15 columns)

- LG전자 서비스 북아메리카 내의 일간 부품 수요 관련 데이터
- 입력 변수
 - 'com_cd'(법인명), 'branch_cd'(창고명)
 - 부품 ID 관련 변수: 'part_no'(부품 ID), div_cd, part_class_cd, division_name, part_group, part_release_date(세부 부품 ID*)
 - 주문 관련 변수: 'ord_type_cd'(주문 종류), 'timestamp'(주문 날짜), 'year', 'week'(해당 주간 집계 연도/주), 'start_date'(주간 집계가 시작된 날짜), 'end_date'(주간 집계가 끝난 날짜)
- 출력 변수
 - 'actual' (부품 주문 개수)

훈련 데이터: 2018/07/27 (금) – 2023/08/04 (금)
경진대회 데이터: 2023/07/31 (월) – 2023/10/29 (일)

1. 일간 데이터 (327743 rows 15 columns)

- LG전자 서비스 북아메리카 내의 일간 부품 수요 관련 데이터
- 입력 변수
 - 부품 ID 관련 변수: div_cd, part_class_cd, division_name, part_group, part_release_date(세부 부품 ID)
 - 세부 부품 ID는 동일한 부품 ID를 가지고 있는 경우에도 다를 수 있음
 - 주문 관련 변수: 'ord_type_cd'(주문 종류)
 - F : 일반 Sales Order
 - W : Warranty 입력 수량 (직영센터)
 - O : OEM 거래선 주문

훈련 데이터: 2018/07/27 (금) – 2023/08/04 (금)
경진대회 데이터: 2023/07/31 (월) – 2023/10/29 (일)

2. 주간 데이터 (870525 rows 6 columns)

- 일간 데이터를 주간으로 변경 (한 주는 월요일부터 일요일로 구성)
- 수요가 0인 주간을 나타내는 데이터 추가
- 입력 변수
 - 'com_cd'(법인명), 'branch_cd'(창고명)
 - 부품 ID 관련 변수: 'part_no'(부품 ID)
 - 주문 관련 변수: 'week', 'year'(해당 주간 집계 연도/주), 'start_date'(주간 집계가 시작된 날짜), 'end_date'(주간 집계가 끝난 날짜)
- 출력 변수
 - 'actual' (부품별 주간 총 주문 개수)

훈련 데이터: 2018/07/27 (금) – 2023/08/04 (금)
경진대회 데이터: 2023/07/31 (월) – 2023/10/29 (일)

3. 부품 별 비용 (3371 rows 4 columns)

- Shipping order cost: 주문 부품 개수별 발생
- Airplane order cost: 주문을 적게 한 경우 발생
- Carrying cost: 주문을 많이 한 경우 발생

part_no	order_cost_air	order_cost_ship	carrying_cost
TEST_0	9.081425567	2.016744433	0.13835567
TEST_1	9.081425567	2.016744433	0.13835567
TEST_10	98.43405824	21.85960095	1.49964452
TEST_100	46.6944024	10.3695918	0.71139
TEST_1000	48.13409577	10.68930962	0.733323753
TEST_1001	63.7591389	14.15921845	0.971371545
TEST_1003	44.65316835	9.916287705	0.680291765
TEST_1005	267.9401274	59.50241583	4.082072314
TEST_1006	190.1510582	42.22752091	2.896954545
TEST_1007	49.215568	10.929476	0.7498
TEST_1012	234.8924836	52.16340818	3.578590909

훈련 데이터: 2018/07/27 (금) – 2023/08/04 (금)
경진대회 데이터: 2023/07/31 (월) – 2023/10/29 (일)

4. 제출용 데이터

- 3371개중 선정된 10개의 부품별 주간 주문량
- 실제 예측 기간은 2023/07/31부터 2023/10/29
- 양식상 예측 기간은 07/31부터지만 08/04의 데이터까지 활용 가능

part_no	year	week	Task 1	Task 2	start_date	end_date
TEST_145	2023	31			2023-07-31	2023-08-06
TEST_145	2023	32			2023-08-07	2023-08-13
TEST_145	2023	33			2023-08-14	2023-08-20
TEST_145	2023	34			2023-08-21	2023-08-27
TEST_145	2023	35			2023-08-28	2023-09-03
TEST_145	2023	36			2023-09-04	2023-09-10
TEST_145	2023	37			2023-09-11	2023-09-17
TEST_145	2023	38			2023-09-18	2023-09-24
TEST_145	2023	39			2023-09-25	2023-10-01
TEST_145	2023	40			2023-10-02	2023-10-08
TEST_145	2023	41			2023-10-09	2023-10-15
TEST_145	2023	42			2023-10-16	2023-10-22
TEST_145	2023	43			2023-10-23	2023-10-29
TEST_4290	2023	31			2023-07-31	2023-08-06

1. 결과는 제출용 데이터 파일의 'Task1', 'Task2'란을 채워 제출

part_no	year	week	Task 1	Task 2	start_date	end_date
TEST_145	2023	31			2023-07-31	2023-08-06
TEST_145	2023	32			2023-08-07	2023-08-13
TEST_145	2023	33			2023-08-14	2023-08-20
TEST_145	2023	34			2023-08-21	2023-08-27
TEST_145	2023	35			2023-08-28	2023-09-03
TEST_145	2023	36			2023-09-04	2023-09-10
TEST_145	2023	37			2023-09-11	2023-09-17
TEST_145	2023	38			2023-09-18	2023-09-24
TEST_145	2023	39			2023-09-25	2023-10-01
TEST_145	2023	40			2023-10-02	2023-10-08
TEST_145	2023	41			2023-10-09	2023-10-15
TEST_145	2023	42			2023-10-16	2023-10-22
TEST_145	2023	43			2023-10-23	2023-10-29
TEST_4290	2023	31			2023-07-31	2023-08-06

2. 최종 결과는 non-negative integer로 변형하여 제출

- 변형하지 않는 경우 발생하는 불이익은 제출자 책임

3. 테스트 기간 주문량 예측 시 2023년 08월 04일 이전까지 관측된 외부 데이터만 활용 가능
- 현재 시점(2023년 10월)에는 테스트 기간 데이터 (2023년 8월 이후) 동안에 관측된 외부데이터를 수집할 수 있음
 - 하지만 **예측시점은 2023년 8월 4일**이기때문에 테스트 기간의 데이터를 활용할 수 없음.

추가 문의사항은 오픈채팅방 활용
(<https://open.kakao.com/o/gecuDcOf>
참여코드: DSC3)

Thank you!

공동주최

POSTECH UNIST KAIST

주관

ime 포항공과대학교
산업경영공학부

후원

LG전자 ECMiner Hankook & Company POSTECH HOLDINGS PuzzleData