InstaCart machinelearning20220527

```
 |-- up_cnt: long (nullable = true)
 |-- up_reord_cnt: long (nullable = true)
 |-- up_no_reord_cnt: long (nullable = true)
 |-- up_reoredered_avg: double (nullable = false)
 |-- up_max_ord_num: integer (nullable = true)
 |-- up_min_ord_num: integer (nullable = true)
 |-- up_avg_cart: double (nullable = false)
 |-- up_avg_prior_days: double (nullable = false)
 |-- up_max_prior_days: double (nullable = false)
 |-- up_min_prior_days: double (nullable = false)
 |-- up_avg_ord_dow: double (nullable = false)
 |-- up_avg_ord_hour: double (nullable = false)
 |-- up_usr_ratio: double (nullable = false)
 |-- up_usr_reord_ratio: double (nullable = false)
 |-- up_usr_ord_num_diff: integer (nullable = true)
 |-- usr_total_cnt: long (nullable = true)
 |-- prd_uq_cnt: long (nullable = true)
```

Command took 0.03 seconds -- by kjr5189@gmail.com at 2022. 6. 4. 오전 1:37:59 on InstaCart

Cmd 92

```
# 여러 값으로 구성된 vector 컬럼에서 특정 값만 추출. probability 컬럼은 0/1 일때의 확률을 모두 가짐. 이중 1일 때(즉 재주문)의 확률을 추출
# 먼저 vector를 array로 변환
from pyspark.ml.functions import vector_to_array
predictions = predictions.withColumn("probability_arr", vector_to_array('probability'))
display(predictions.limit(10))
```

▸ (3) Spark Jobs

	user_id	product_id	order_id	up_cnt	up_reord_cnt
1	172302	19822	1243888	1	0
2	1623	43908	1778015	2	1
3	160722	46979	2741763	11	10
4	137485	21903	1392405	5	4