

## 주요 경력기술서

### [타겟 마케팅 용 base table ETL 구축]

목표	- Category 별 Propensity 모델(PP)과 Incremental value regression 모델(NBA)을 활용해 각 고객별 타겟 유망성 base table 구축
사용기술	- Airflow, Hive
과정	- PP 와 NBA 간 카테고리를 재정의하고 Mapping 시켜 Join - Key metric 이 Incrementality 이기 때문에 NBA 와 PP 점수가 동시에 높은 Core 고객, 그리고 아닌 고객으로 구분 - 고객별로 해당 카테고리 별 ranking, purchase type, membership 정보를 Daily batch 실행
성과	- 기존 PP 모델만 쓸 때 보다 Incrementality 가 평균 23% 상승 - 캠페인별 유망 고객 추출 시 일관성 확보 - 카테고리 별 동일한 컨트롤 그룹으로 성과 오염 방지 - 고객의 Daily 상태 값으로 최신성 확보

### [타겟 고객 추출 자동화]

목표	- in-app 타겟 마케팅 별 유망 고객 추출 자동화 - 유망 고객 : Incrementality revenue 를 최대화할 수 있는 고객
사용기술	- Airflow, Spark, Hive, Jira
과정	- Jira 를 통해서 Marketing planner 가 캠페인 데이터 입력 - NBA&PP base table 기반 코어고객을 우선순위로 추출 - Campaign Manager 와 추출된 Target 을 연동하기 위한 S3 공간 구축 - Airflow Schedule 에 따라 Daily 로 고객 자동 추출
성과	- Target 추출을 위한 Data 분석가 Workforce 를 zero 화 - 현재 전체 캠페인 중 자동추출이 30 ~ 40% 차지 - 미국 특허 신청 완료

### [Audience Connector 와 타겟 고객 파일 연동 자동화]

목표	- 캠페인 세팅을 위해 타겟이 포함된 csv 파일을 Audience Connector 에 수동으로 업로드하는 부분을 자동화
사용기술	- Airflow, S3, Spark
과정	- Target 이 포함된 변수를 Spark 를 통해 S3 상에 csv 로 Write - Audience Connector 개발팀에서 Target 이 업로드 된 S3 에 접근하여 1 시간에 1 번씩 Sync 작업으로 파일 복사 실행
성과	- 매주 캠페인 타겟 세팅을 위해 소비되는 시간 및 인력 zero 화 - 주 평균 250 개 csv file 생성 및 전달 작업 자동화

[Personalized re-targeting]

목표	- ATC 또는 제품 상세 페이지를 조회하거나 검색을 한 고객을 Re-targeting 하기 위한 Daily-based ETL 구축
사용기술	- Airflow, Hive
과정	- ATC 또는 제품 상세페이지를 조회한 고객을 log schema 를 통해 추출 - 해당 제품 또는 유사 제품을 구매 또는 노출할 경우 등 제외 로직 적용 - A/B test 결과 구매율에 대한 통계적 유의성 확인
성과	- Weekly incrementality metric 기준 10 억 달성 (vs. control 그룹)

[Time-series regression model 개발]

목표	- 국가별 전략 스마트폰 (Galaxy S, Galaxy Note) 판매량 예측
사용기술	- R, Python, Postgres, AWS
과정	- STL 기반 Seasonality 와 Trend decomposition 을 통해 data transformation - Covid 로 인한 단기 Trend 와 장기 Trend 를 분리하여 recovering trend 반영 - Seasonality 를 Group lasso 모델을 통해 예측 - AWS 로 DB migration 을 위해 R → Python 으로 코드 전환
성과	- 정확도(weighted sum average) 75 ~ 83% 달성하여 SCM 안정화 기여 - Covid 로 인한 예측 불확실성을 이슈 해결

[Markov chain attribution model 개발]

목표	- Paid marketing 채널 별 기여도 측정
사용기술	- Markov Chain, Graph theory, R
과정	- 기존 First touch 또는 Last touch 기반에 측정된 Attribution (= budget/conversion)에서 Markov Chain 기반으로 모든 채널 Path 에 따른 기여도 측정 - Web(samsung.com)방문자 Path 를 바탕으로 각 채널 별 Transition probability 를 계산하고, 이를 매트릭스로 전환 - Markov Chain 이론으로 steady state 통해 각 채널 별 구매비중을 계산 - 총 구매자 수와 조합을 하며 각 채널 별 Ad cost per Acquisition 측정
성과	- First touch 시 Facebook, Last touch 시 Google search 채널의 효과가 좋지만 Attribution model 기반 Search > Facebook > Banner 기여도 확인 - Non-linear optimization 을 통해서 채널 별 Budget 을 Facebook: Search: Banner = 4:4:2 로 조정

[ML-based Propensity model 개발]

목표	<ul style="list-style-type: none"> <li>- samsung.com 재방문 확률이 높은 고객을 찾음</li> <li>- 해당 고객의 Cookie 를 기반으로 Paid channel 을 통한 Retargeting 실행</li> </ul>
사용기술	<ul style="list-style-type: none"> <li>- SparkML, Hive</li> </ul>
과정	<ul style="list-style-type: none"> <li>- Web log 데이터에서 최근 90 이내 방문자 중 미래 30 일동안 방문하면 1 아니면 0 이라는 Binary classification 모델 개발</li> <li>- Web log 기반 feature 를 25 개 생성</li> <li>- Airflow 로 해당 feature 를 주기적으로 테이블에 저장</li> </ul>
성과	<ul style="list-style-type: none"> <li>- Recall 기준 60 ~ 65% 수준</li> <li>- 국가별 평균 CTR 2 배 상승</li> </ul>