Ch 11. Case Studies

실전 케이스 스터디 — End-to-End 문제 해결

이 챕터에서 다루는 것

케이스 인터뷰는 **제품 감각 + 통계 + ML + 시스템 설계**를 결합한 종합 문제임

- 단독 SQL이나 통계 문제가 아니라 **비즈니스 문제를 처음부터 끝까지** 해결
- 회사마다 온사이트 면접 또는 **Take-Home 프로젝트** 형태로 출제
- Ch 7(ML) + Ch 10(Product Sense)의 교차점에 해당
- 이 장에서 **8개 실전 케이스**(Facebook, Amazon, Stripe 등)를 풀어봄

Part 1

케이스 인터뷰 접근법

케이스 인터뷰 성공을 위한 5가지 팁

Clarify	Be Coachable	Pragmatic	Timebox
비즈니스/제품 목표를 먼저 확인. 가장 흔한 실수는 목표 확인 없이 솔루션부터 제시하는 것	면접관의 힌트를 적극 수용. "좋은 포인트네요, XGBoost도 시도해볼 수 있겠네요"	면접관 팀이 수개월 작업한 문제 — 혁신적 솔루션보다 **합리적이고 실현 가능한** 접근 우선	토끼굴(rabbit hole) 금지. 초안 답변 후 "더 깊이 들어갈까요?" 물어볼 것

Take-Home 챌린지의 핵심: 스토리텔링

Take-Home은 케이스 인터뷰와 유사하나 **데이터셋이 제공**된다는 점이 다름

- 가장 중요한 것: **분석 결과를 이야기(narrative)로 전달**하는 능력
- 숫자와 시각화를 스토리에 엮어 **의사결정자가 이해할 수 있게** 전환
- 채용 매니저가 가장 높이 평가하는 역량 = **분석 → 접근 가능한 추천**(recommendation)
- 기술적 정확성뿐 아니라 **비기술 이해관계자를 설득**할 수 있어야 함

케이스 인터뷰에서 트레이드오프 언급이 중요한 이유

단일 정답은 없음 — **대안과 한계를 함께 제시**하면 깊이를 보여줌

측면	예시
모델 복잡도	선형회귀(해석 용이) vs. 신경망(정확도 높으나 과적합 위험)
데이터 품질	1st-party(정확하나 적음) vs. 3rd-party(풍부하나 편향 가능)
배포 전략	일괄 배포(단순) vs. A/B 테스트 배포(안전하나 복잡)
지표 선택	Precision(FP 최소화) vs. Recall(FN 최소화)
비용	자동 판단(빠르나 오류 있음) vs. 인간 리뷰(정확하나 느림)

Part 2

매출 추정 & 비즈니스 모델링

Case 11.1 — Citadel: 소매 매출 예측

문제: 미국 상장 소매 체인의 실물 매장 매출을 추정하라. GPS 기반 **유동인구(foot traffic) 데이터**(1천만 대 모바일)를 활용

- **목적**: 분기 실적 발표 전에 매출을 예측하여 **투자 결정**(long/short)에 활용
- **데이터**: 일별 date-storeID 쌍, 방문 횟수(비정규화 정수)
- **핵심 접근**: foot traffic을 분기 단위로 롤업 → **정규화** → 단순 회귀 모델
- **과적합 방지**: 데이터가 12분기뿐 → 유사 소매업체(Walmart, Costco) 데이터 합산

Citadel: 패널 정규화와 샘플링 편향

유동인구는 **표본 패널**에서 나온 데이터이므로 보정이 필수임

- **정규화 공식**: (표본 방문수 / 표본 인구) × 전체 인구
  - 지역별(zip code, census block)로 세분화하면 더 정확함
- **샘플링 편향 원인**: 스마트폰 미보유, 위치 앱 미설치, GPS 권한 비활성화
- **편향 검증**: 패널 기기의 야간 위치 → 거주지 추정 → Census 데이터와 인구분포 비교
- **회귀 문제**: 피처 > 데이터 포인트 → PCA 차원축소, 다중공선성 → VIF 확인

Citadel: 제3자 데이터의 한계

foot traffic 데이터는 강력하나 **외부 의존성 리스크**가 있음

- 앱 패널 구성이 비공개이고 수시로 변동 → **일관된 백테스트가 어려움**
- 방문 = 매출이 아님 — 매장을 방문해도 **구매하지 않을 수** 있음
- 방문 판정 자체가 지오펜스(geofence) 기반 **모델링 결과**이므로 오류 가능
- **보완**: 영수증(receipts), 신용카드, POS 데이터와 교차 검증

Part 3

추천 시스템

Case 11.2 — Amazon: Prime Video 추천

문제: Amazon Prime Video에서 사용자에게 어떤 프로그램을 추천할 것인가?

- **핵심 기법**: Collaborative Filtering — "비슷한 사용자가 좋아한 콘텐츠"를 추천
- **데이터**: m(사용자) × n(프로그램) 매트릭스, 각 셀 = 평점(1~5) 또는 미시청
- **유사도 측정**: 사용자 행(row) 간 **코사인 유사도**. 사용자별 평점 편향은 정규화로 보정
- **추천 로직**: 유사 사용자가 높게 평가 + 본인이 미시청인 프로그램 → 가중평균 순위

Amazon: 콜드 스타트와 대안 기법

콜드 스타트 해결

- **신규 프로그램**: 별도 "New to Amazon" 패널로 노출. 장르·배우·언어 등 콘텐츠 메타데이터로 유사도 계산
- **신규 사용자**: 인구통계 기반 유사도 산출. 데이터 축적 전까지 **인기/트렌드** 콘텐츠 추천

A/B 테스트 주의점

- 상위 지표(total watch time) p=0.04로 유의 → **바로 배포하면 안 됨**
- 실험 최소 2주 이상 유지, 요일 효과 고려
- **반대 지표**(counter metric) 확인: 추천 정밀도, 추천 vs. 자발 시청 평점
- 통계적 유의 ≠ 실질적 유의(practical significance)

Case 11.8 — Instagram Explore 추천

문제: 팔로우하지 않는 계정의 콘텐츠를 개인화 추천. 월 10억 MAU, 분당 100만 피드 새로고침 규모

- **핵심 전략**: 콘텐츠(media) 레벨이 아닌 **계정(account) 레벨** 협업 필터링
  - 콘텐츠 우주가 너무 크고 매초 신규 생성 → 계정 단위로 후보 축소(candidate retrieval)
- **임베딩**: 사용자가 상호작용한 계정 시퀀스를 word2vec 방식으로 학습 ("ig2vec")
- **유사 계정 검색**: 코사인 거리 / dot product → KNN으로 후보군 생성

Instagram Explore: 랭킹과 배포 전략

후보 축소(retrieval) 후에는 **개인화 랭킹**으로 최종 순서를 결정함

- **랭킹 모델**: 다중분류 신경망 — 각 포스트에 대해 like / comment / share / hide / report 확률 예측
  - 각 행동에 가중치 부여 → 최종 추천 점수 산출
- **온라인 추론**: 실시간 피처를 Cassandra 등에 사전 저장, 저지연 서빙
- **배포 전략**: 4가지 중 **A/B 테스트 배포(flighting)** 가 가장 적합
- **모델 드리프트**: KL divergence로 피처 분포 변화 모니터링 → 주기적 재학습

Instagram Explore: A/B 테스트에서 지표가 안 움직이는 이유

모델은 개선되었는데 비즈니스 지표(engagement)와 무상관일 수 있음

- **지표 포화**: 콘텐츠 관련성이 이미 충분 → 추가 모델 개선이 인게이지먼트에 반영 안 됨
- **과도한 개인화**: 추천이 너무 정확하면 사용자가 **오히려 불쾌감**(creepy) 느낄 수 있음
- **교훈**: 모델 정확도 ↑ ≠ 사용자 만족 ↑ — 최적점(sweet spot)이 존재함

Part 4

가격 최적화 & 수익 모델링

Case 11.4 — Walmart: 상품 가격 최적화

문제: 북미 5,000개 매장 × 100,000개 상품 = **5억 개 가격** 결정 알고리즘 구축

- **비즈니스 목표**: 이익(profit) 극대화. 단, "Everyday Low Price" 브랜드 전략과 균형
- **업데이트 빈도**: 잦은 변경 → 소비 변화 대응 가능 / 물리적 라벨 교체 비용 + 브랜드 훼손
  - 절충안: 초기가 → 할인가 → 최종 클리어런스의 **3단계 가격**
- **핵심 접근**: 개별 상품마다 **수요 곡선(demand curve)** 구축

Walmart: 수요 곡선과 가격 탄력성

상품별 선형 회귀로 가격-수요 관계를 모델링함

- **이익 함수**: (판매가 - 원가) × 수요량, 수요 = f(가격 변화)
- **가격 탄력성**: 회귀 계수의 부호와 크기로 해석
  - 큰 음수(탄력적): 가격↑ → 수요 급락 (예: 맥주 — 와인/하드셀처로 대체 가능)
  - 작은 음수(비탄력적): 가격 변화에 둔감 (예: 담배, 처방약)
  - 양수(Veblen재): 가격↑ → 수요↑ (예: 프리미엄 와인)
- **검증**: 유사 상품군(홈에센셜 등)끼리 탄력성이 비슷한지 확인

Walmart: 블랙박스 모델과 A/B 테스트

추가 피처

- **상품**: 조달 원가, 진열 위치, 카테고리, 용량
- **경쟁사**: 온/오프라인 경쟁 가격
- **재고**: 현재 수량, 다음 입고일, 시즌/유통기한
- **이력**: 과거 판매량 및 가격 이력

A/B 테스트 설계

- 가격 탄력성·계절성이 유사한 2개 카테고리 선택
- 장바구니 교차 효과 방지: 게임기 ↔️ 게임 소프트 조합은 금지
- 예: 치약(Control) vs. 화장지(Test)
- 수개월 관찰 → 매출/이익/재고 회전율 측정

Case 11.3 — Airbnb: 신규 숙소 연간 수익 예측

문제: 신규 등록 숙소의 연간 수익을 예측하여 잠재 호스트에게 보여주기

- **피처**: 침실/욕실 수, 면적, 야간 요금, 최소 숙박일, zip code, 주변 명소 거리, 인근 숙소 가격·점유율
- **결측 처리**: 면적(square footage) 10% 누락 → 평균 대치는 위험
  - 누락 자체가 신호일 수 있음(면적이 좁아서 비공개)
  - 다른 피처(침실 수, 욕실 수)로 면적을 **예측하는 보조 모델** 구축
  - 또는 **제3자 데이터**(county parcel records) 매칭
- **모델**: Occam's Razor → 선형회귀 베이스라인 → 랜덤 포레스트/XGBoost 확장

Airbnb: 고차원 피처와 모델 선택 트레이드오프

피처 수백 개가 되면 **차원의 저주**가 발생함

전략	방법	적합 상황
Feature Selection	상관관계, VIF, RFE	피처 간 관계 파악 가능할 때
PCA / SVD	분산 기반 차원 축소	해석보다 예측 정확도 우선 시
임베딩 / 오토인코더	딥러닝 기반 축소	데이터 충분 + 비선형 관계

- **모델 복잡도**: 선형회귀(해석↑) → RF/XGBoost(균형) → 신경망(정확도↑, 과적합 위험)
- **QPS 요구사항**: 높은 트래픽이면 단순 모델이 레이턴시에 유리 → 정확도-비용 트레이드오프

Part 5

텍스트 분석 & 감성 분석

Case 11.5 — Accenture: 호텔 브랜드 소셜 리스닝

문제: 주요 호텔 체인의 Facebook/Twitter/Reddit 상 브랜드 언급을 분석하라

- **왜 중요한가**: Yelp/TripAdvisor 리뷰만으로는 부족 — SNS의 바이럴 효과가 더 큼
- **전략적 가치**: 브랜드 인식 모니터링 + 부정 감성의 공통 원인 파악 → NPS 개선
- **실행 가능한 조치**:
  - 부정 게시물 작성자에게 **선제적 환불/보상** 제안
  - 문제 숙소에 **실시간 알림** → 현장 즉시 대응
  - 원인 분석 후 서비스 프로세스 수정

호텔: 감성 분석과 토픽 분류 파이프라인

감성 분석(Sentiment)

- **전처리**: 인코딩 수정, HTML 제거, 불용어 제거, 어간 추출, 벡터화
- **텍스트 벡터화**: BoW, N-gram, TF-IDF
- **모델**: Logistic Regression, SVM, 신경망 등 분류기
- **평가**: 혼동 행렬 + Precision / Recall

토픽 분류(Topic)

- 리뷰를 체크인, 객실, 룸서비스 등 **주제별 자동 분류** → 담당 부서로 라우팅
- **방법**: 벡터 기반 K-means 클러스터링 또는 **LDA**(Latent Dirichlet Allocation)
- LDA: 각 게시물 = 토픽 분포, 각 토픽 = 단어 분포로 모델링

Part 6

소셜 그래프 & 친구 추천

Case 11.6 — Facebook: People You May Know (PYMK)

문제: "알 수도 있는 사람" 친구 추천 기능 구축. 목표 = 의미 있는 연결(meaningful connections) 증가

- **접근 1**: 연락처 업로드 기반 — 주소록의 Facebook 사용자를 추천
- **접근 2**: 소셜 그래프 기반 — 2차·3차 연결 중 **친구 가능성** 높은 순서로 랭킹
- 실제로는 두 접근을 **혼합(blend)** 하여 사용

PYMK: 피처와 모델

프로필 유사도

나이, 모교, 직장, 고향, 현재 도시, 공통 친구 수

앱 내 활동

동일 친구와 높은 상호작용, 같은 이벤트 참석, 프로필 방문, 같은 글에 댓글, 함께 태그된 사진

생태계 시그널

Instagram 팔로우, Messenger/WhatsApp/DM 대화

외부 시그널

이메일, 전화번호, 모바일 GPS 위치 근접성

PYMK: 콜드 스타트와 신규 사용자 전략

Facebook은 친구가 없으면 피드가 비어있는 **콜드 스타트 문제**가 심각함

- 신규 사용자 → 연락처 미업로드 시 추천할 데이터 부족
- **Facebook의 전략적 이유**: Reddit/YouTube와 달리 친구 콘텐츠 없이는 가치 제공 불가
- **제품 아이디어**:
  - 신규 사용자를 기존 사용자 PYMK에 **부스트**하여 인바운드 요청 증가
  - 친구 수 도달 전까지 **PYMK 노출 빈도 증가**
  - **게이미피케이션**: 프로필 완성 + 친구 추가 진행률 바
  - 친구 수락 직후 비지 않은 PYMK 패널로 **연쇄 친구 추가** 유도

Part 7

대출 승인 & 리스크 모델링

Case 11.7 — Stripe: 소기업 대출 승인 모델

문제: 소기업 대출 신청을 승인/거절하는 모델. 이진 분류 — 전액 상환 or 디폴트

- **모델 출력**: 디폴트 확률 스코어 → 임계값(예: 0.5)으로 이진 판정
- **평가 지표**: Precision-Recall 곡선 (Accuracy 부적합 — 클래스 불균형 심함)
- **FP**: 상환할 대출을 거절 → 이자 수익 10% 기회손실
- **FN**: 디폴트할 대출을 승인 → **원금 전체 손실**(write-off)
- **비용 비율**: FN : FP = **10 : 1** → 가중 정밀도/재현율로 임계값 최적화

Stripe: 피처, 이상 탐지, 인간 리뷰

대출 모델은 **두 가지 피처 차원**을 사용함

차원	피처 예시
신청자 수준	인구통계, IP, 브라우저, 은행 잔고, 신용도, 미결 유치권
신청 수준	질문 완성도, 대출 금액, 용도, 담보 규모

- Reject Inference: 거절된 신청 데이터도 별도 모델로 학습 → 샘플링 편향 보정
- **이상 탐지**: 과거 디폴트 패턴이 변할 수 있으므로 실시간 이상 탐지 보완
- **경계선 사례**: 임계값 근처 → **인간 리뷰**(tiered system)로 FP/FN 추가 감소

Stripe: A/B 테스트와 p-value 판단

대출의 특수성 — 상환 기간이 길어 A/B 테스트가 어려움

- **오프라인 비교**: 동일 데이터에서 모델별 지표 비교 + paired t-test
- **p = 0.06이면?**: 0.05를 약간 넘었다고 즉시 기각하지 말 것
  - 더 많은 데이터로 실험 연장 → p-value 변동 관찰
  - **효과 크기**(effect size)와 반대 지표(counter metrics) 함께 평가
  - 런칭 결정은 p-value 하나가 아닌 **종합적 증거**로 판단

Part 8

케이스별 핵심 패턴 종합

8개 케이스에서 반복되는 공통 패턴

1. **항상 명확화(Clarify)부터**: 모든 케이스의 첫 발화는 "비즈니스 목표가 뭔가요?"
2. **데이터 이해 → 전처리 → 모델링**: 데이터 특성(결측, 편향, 스케일)을 먼저 파악
3. **단순 모델 → 복잡 모델**: Occam's Razor — 선형회귀 베이스라인 후 확장
4. **지표 선택은 비즈니스와 연결**: Accuracy보다 Precision/Recall, 비용 비율 반영
5. **A/B 테스트로 검증**: 단, 반대 지표 + 통계적·실질적 유의성 모두 확인
6. **콜드 스타트 & 편향 대응**: 신규 사용자/데이터, 샘플링 편향을 반드시 언급
7. **트레이드오프 명시**: 복잡도, 비용, 해석성, 레이턴시 간 균형

케이스 유형별 핵심 기법 정리

추정 & 가격

- foot traffic → 회귀
- 수요 곡선 + 가격 탄력성
- 패널 정규화, 편향 검증
- Census 데이터 교차 검증

추천 시스템

- Collaborative Filtering
- 콘텐츠 기반 필터링
- ig2vec 임베딩, KNN
- 콜드 스타트 해결 전략

리스크 & 분류

- 비용 비대칭 (FN >> FP)
- Reject Inference
- 인간 리뷰 티어 시스템
- 감성 분석 + 토픽 분류

케이스 인터뷰 체크리스트

면접 전에 이 흐름을 내재화하면 어떤 케이스든 구조적으로 접근할 수 있음

단계	질문	예시
1. 명확화	비즈니스 목표? 데이터 형태?	"투자 결정용인가요, 사용자 노출용인가요?"
2. 데이터	어떤 피처? 결측/편향은?	"면적 10% 누락 — 보조 모델로 추정"
3. 모델	베이스라인 → 확장?	"선형회귀 → XGBoost → 신경망"
4. 평가	어떤 지표? 비용 구조?	"FN이 FP보다 10배 비쌈"
5. 배포	A/B? 롤아웃? 재학습?	"flighting 배포 + KL divergence 모니터링"
6. 한계	트레이드오프? 엣지 케이스?	"추천이 너무 정확하면 사용자가 불쾌"